EuroMov Digital Health in Motion

Multimodal Analysis of Spontaneous Gestures and Speech in Dyadic Interaction

Matys Precloux [1], Patrice Guyot [1] , Ludovic Marin [1]

[1] EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Montpellier, France

1 INTRODUCTION

Although human communication relies on both verbal and non-verbal cues (such as gestures, posture, and body movements), the latter play a more important role (Schmidt & Richardson, 2008). This study explores which postural condition best supports spontaneous non-verbal moves during a human communication.

Our study is part of the ANR-funded **SYNCOGEST** project, which models spontaneous gestures in face-to-face interactions in order to enhance (in mid-term) embodied conversational agents.

- Human communication relies on gestures **synchronized** with speech.
- The torso, hands, and head convey important interactional information.
- This project aims to **automate** speech-related gestures to facilitate annotation.

•A **preliminary study** will identify which postural condition (sitting, semi-standing, standing) elicits the most spontaneous movement.

•The main protocol will involve multimodal capture (video, audio, motion) of dyadic interactions.

•The goal is to improve the modeling of human-like co-speech gestures for conversational AI.

Schmidt, R. C., & Richardson, M. J. (2008). *Dynamics of Interpersonal Coordination*.

2 MATERIALS & METHODS

Preliminary study

• To determine in **which** condition (1,2,3) spontaneous movement is most prominent

• Total movement per body part will be computed (head, torso, hands), measured using the **MediaPipe** software

• **Comparison** will be made across postural conditions (seated, semi-standing, standing)

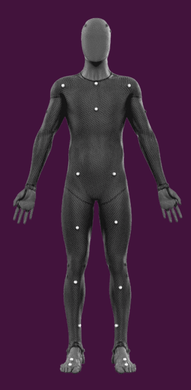
Main study

Participants

• 20 dyads will take part in the study.

• 10 dyads consist of **familiar** interlocutors, and 10 of **unfamiliar** ones.

• Each dyad will engage in a 1-hour face-to-face conversation.



Experimental Setup

• Full-body motion data will be recorded using an **optical motion capture system (Qualisys)**.

• Reflective markers will be placed on key anatomical segments: **head, trunk, arms, and hands**.

Data Collected

• Biomechanical parameters will be extracted, including:

• **Velocity** / **Acceleration** / **Movement trajectories**

$$p = m \cdot \vec{v}$$

$$\vec{a} = \frac{\Delta \vec{v}}{\Delta t}$$

Corpus Annotation

• 25% of the recorded corpus will be **manually annotated**.

• These annotated segments will be used to train and validate an **automatic gesture segmentation model**.

3 HYPOTHESIS

• H1 : People move **more** when standing or semi-standing than when sitting.

• H2 : People move **less** when talking to someone they know than to a stranger

4 FUTURE OUTCOMES

• **Automatic segmentation** model for spontaneous gestures based on multimodal movement features.

• Creation of a labeled gesture database linking **biomechanical** patterns with linguistic function.

• Contribution to the development of realistic gesture generation models for **virtual agents** and **social robots**.